doi:10.6041/j.issn.1000-1298.2023.02.028

基于空间注意力和可变形卷积的无人机田间障碍物检测

杜小强[1,2]　李卓林[1]　马锃宏[1,2]　杨振华[3]　王大帅[4,5]

(1.浙江理工大学机械工程学院,杭州 310018;2.浙江省种植装备技术重点实验室,杭州 310018;
3.龙泉市菇源自动化设备有限公司,龙泉 323700;4.中国科学院深圳先进技术研究院,深圳 518055;
5.广东省机器人与智能系统重点实验室,深圳 518055)

摘要:为了解决植保无人机作业时,传统田间障碍物识别方法依赖人工提取特征,计算耗时较长,难以实现在非结构化田间环境下实时作业识别的问题,提出一种优化的 Mask R-CNN 模型的非结构化农田障碍物实例分割方法。以 ResNet-50 残差网络为基础,将空间注意力(Spatial attention, SA)引入残差结构,聚焦跟踪目标的显著性表观特征并主动抑制噪声等无用特征的影响;引入可变形卷积(Deformable convolution, DCN),通过加入偏移量,增大感受野,提高模型的鲁棒性。构建包含农田典型障碍物的数据集,通过对比试验研究在 ResNet 残差网络结构中的不同阶段中加入空间注意力和可变形卷积时的模型性能差异。结果表明,与 Mask R-CNN 原型网络相比,在 ResNet 的阶段 2、阶段 3、阶段 5 加入空间注意力和可变形卷积后,改进 Mask R-CNN 的边界框(Bbox)和掩膜(Mask)的平均精度均值(mAP)分别从 64.5%、56.9% 提高到 71.3%、62.3%。本文提出的改进 Mask R-CNN 可以很好地实现农田障碍物检测,可为植保无人机在非结构化农田环境下安全高效工作提供技术支撑。

关键词:田间障碍物;Mask R-CNN;空间注意力;可变形卷积

中图分类号:TP391.4　　**文献标识码**:A　　**文章编号**:1000-1298(2023)02-0275-09　　**OSID**:

UAV Field Obstacle Detection Based on Spatial Attention and Deformable Convolution

DU Xiaoqiang[1,2]　LI Zhuolin[1]　MA Zenghong[1,2]　YANG Zhenhua[3]　WANG Dashuai[4,5]

(1. *School of Mechanical Engineering, Zhejiang Sci-Tech University, Hangzhou* 310018, *China*
2. *Key Laboratory of Transplanting Equipment and Technology of Zhejiang Province, Hangzhou* 310018, *China*
3. *Longquan Guyuan Automation Equipment Co., Ltd., Longquan* 323700, *China*
4. *Shenzhen Institute of Advanced Technology, Chinese Academy of Sciences, Shenzhen* 518055, *China*
5. *Guangdong Provincial Key Laboratory of Robotics and Intelligent System, Shenzhen* 518055, *China*)

Abstract: In order to solve the problem that the traditional field obstacle recognition methods rely on manual feature extraction, long calculation time, and it's difficult to achieve real-time recognition in unstructured field environment, an optimized unstructured field obstacle instance segmentation method based on Mask R-CNN model was proposed. Firstly, an unstructured field obstacle dataset was constructed by aerial photography and network search. And then based on the ResNet-50 residual network, the spatial attention was introduced to focus on the significant apparent features of the tracking target, and the influence of useless features such as noise was suppressed. In addition, the deformable convolution was introduced into the structure of the ResNet-50 to add the offset, increase the receptive field and improve the robustness of the model. Comparative analysis was made by adding spatial attention and deformable convolution to different stages in the structure of ResNet-50. The results showed that compared with the original Mask R-CNN model, the mAP values of Bbox and Mask in Mask R-CNN improved by adding spatial attention and deformable convolution in Stage 2, Stage 3 and Stage 5 of the ResNet-50 were increased from 64.5% and 56.9% to 71.3% and 62.3%, respectively. The improved

收稿日期:2022-03-14　修回日期:2022-06-01

基金项目:国家自然科学基金项目(32001424、31971798)、深圳市科技计划项目(JCYJ20210324102401005)、国家重点研发计划项目(2022YFD2202103)、浙江省"领雁"研发攻关计划项目(2022C02057)和浙江省"三农九方"科技协作计划项目(2022SNJF017)

作者简介:杜小强(1978—),男,教授,博士,主要从事农业机器人研究,E-mail:xqiangdu@zstu.edu.cn

通信作者:王大帅(1990—),男,高级工程师,博士,主要从事人工智能与精准农业研究,E-mail:ds.wangl@siat.ac.cn

Mask R-CNN can well realize field obstacle detection and provide technical support for plant protection UAV to work safely and efficiently in unstructured field environment.

Key words: obstacle of field; Mask R-CNN; spatial attention; deformable convolution

0 引言

随着机器人技术和人工智能的快速发展,植保无人机逐渐成为我国农业航空产业的重要组成部分[1]。但是我国农田非结构化特点突出,随机离散化分布的障碍物(树木、电线杆、建筑、人、电线塔等)对无人机飞行安全构成严重威胁[2]。准确识别障碍物是无人机进行避障和路径规划的前提,对保证无人机作业效率和飞行安全至关重要。

传统的无人机障碍物检测方法是利用距离探测传感器,如激光雷达[3]、微波传感器[4]、超声波传感器[5]等,感知障碍物的存在。但是,这类方法会受到传感器性能和环境的限制,只能获取有限的距离和轮廓[2]。虽然现今已有研究证明能够通过激光传感器或深度相机等获得的点云直接识别障碍物类别[6],但是由于点云数据的稀疏性,感知障碍物的类别精度较低[7]。为了在 RGB 空间对障碍物进行描述,部分研究人员还研究了基于机器学习算法和单目相机的目标感知方法,但此类方法依赖于人工选取图像特征,计算耗时较长,难以满足无人机高动态、高实时性作业的要求。

近年来,随着人工智能的迅速发展,基于卷积神经网络的深度学习算法在计算机视觉领域展现出强大的性能。随着精准农业的发展,人工智能在其他领域的发展成果开始迁移到农业领域[8-14]。但是深层神经网络计算量大,模型运行速度慢;又由于我国农田环境复杂,非结构化特点突出,随机离散化分布的障碍物会导致障碍物检测困难,直接将 Mask R-CNN 应用于非结构化农田环境下的障碍物检测,会导致模型的精度下降。

MNIH 等[15]最早提出注意力机制。将注意力机制与神经网络结合,将有利于从空间域、通道域深度挖掘图像信息的特征,进而提高神经网络模型的检测精度和速度。黄林生等[16]将多尺度卷积结构和注意力机制结合,提出一种农作物病害识别模型。熊俊涛等[17]在 Deeplab V3 网络的基础上引入稠密特征传递方法和注意力模块,实现在复杂野外环境中为智能疏花提供视觉支持,并且该模型具有较强的鲁棒性和识别率。注意力机制的引入,增强了有用特征的权重,减弱了无用特征的影响,进一步提高了特征提取能力,提高了模型的鲁棒性。

标准卷积的常规采样难以适应目标的形状变化[18],为此,DAI 等[19]提出可变形卷积,替代传统的标准卷积,通过对卷积核中每个采样点位置增加可学习的偏移量,从而增加空间采样位置,可变形卷积核的大小和位置可以根据图像内容发生自适应的变化,从而提高目标检测的精度。SUN 等[20]通过将 RGB 图像与近红外图像融合,并引入可变形卷积对 R-FCN 模型进行改进,解决自然环境中的复杂背景和夜间光线不足造成甜菜幼苗和杂草识别困难的问题。可变形卷积的引入提高了网络对图形几何变形的适应能力,进而提高模型的特征提取能力。

我国非结构化农田环境中随机离散分布的障碍物对植保无人机的飞行安全和作业效率有直接影响。传统图像识别方法依赖人工提取特征,计算耗时较长,难以适应非结构化田间复杂环境下的实时作业要求。深度学习算法虽然在图像分类、目标检测和图像分割等领域应用广泛,但在农田障碍物检测中的应用尚有不足。

本文基于空间注意力机制和可变形卷积对 Mask R-CNN 模型进行优化,解决现有的深度学习模型对田间障碍物的检测精度低、鲁棒性较差等问题。

1 数据集构建

在文献[21]的研究基础上,通过无人机航拍、手持相机拍摄和网络搜索等方法,采集多环境、多场景、多视角下的田间典型障碍物图像信息,对文中数据集进行补充,包括树木、电线杆、建筑、电线塔、无人机、人共 6 类障碍物,一共 6 000 幅图像。同时,为了减少计算量,降低模型训练时间,将原图像调整为 416 像素×416 像素。随后,用 Labelme 图像标注工具标注出障碍物图像轮廓,共标注目标 11 578 个,制作成 COCO 格式的数据集。在数据集中随机选取 4 800 幅图像作为训练集,600 幅图像作为验证集,600 幅图像作为测试集,比例为 8∶1∶1。图 1 为 6 类障碍物图像。

2 田间障碍物实例分割模型

Mask R-CNN 是一种先进的实例分割算法,具有目标检测和实例分割两大功能,能够精确地检测目标并准确地分割目标,在性能上超过了 Faster R-CNN,是一种综合性能优异的实例分割算法。Mask R-CNN 是一个两阶段的框架,第 1 阶段是通过主

图 1　田间障碍物图像示例

Fig. 1　Examples of field obstacle images

干网络(残差神经网络(ResNet)和特征金字塔网络(Feature pyramid network, FPN))提取图像特征,并通过区域生成网络生成感兴趣区域;第 2 阶段用于分类提议区域并生成边界框和掩膜。

针对非结构化农田障碍物的特点,对现有的 Mask R-CNN 实例分割网络进行改进,构建一种适用于田间障碍物图像检测和分割的网络。本文主要对主干网络做出以下改进:在 ResNet 网络的阶段 2、阶段 3、阶段 5 加入空间注意力机制和可变形卷积。

2.1　基础网络选取

在计算机图像视觉里,卷积神经网络的网络层数越深,能获取到的信息就越多,特征也就越丰富。但是随着网络层数的不断加深,就会出现梯度消失或梯度爆炸的问题[22],导致优化效果更差,测试数据和训练数据的准确率降低。针对这个问题,对输入层和中间层进行归一化操作,这可以使得具有数十层的网络能够开始用反向传播进行随机梯度下降(SGD),从而让网络达到收敛。然而当更深层次网络开始收敛时,出现网络退化问题,增加层数却导致更大的误差。为解决这个问题,HE 等[23]提出了残差网络。残差网络的核心在于 ResNet 残差块结构。

ResNet 残差块使用 Shortcut connection(捷径连接)的连接方式进行 Identity mapping(恒等映射),将输入 $\boldsymbol{x}$ 与经过堆叠的权重层得到的 $\boldsymbol{F}(\boldsymbol{x})$ 进行跨层连接,既不增加额外参数,也不会增加计算复杂性。当 $\boldsymbol{x}$ 和 $\boldsymbol{F}$ 维度相同时有

$$\boldsymbol{y}=\boldsymbol{F}(\boldsymbol{x},\{\boldsymbol{W}_i\})+\boldsymbol{x} \tag{1}$$

其中

$$\boldsymbol{F}=\boldsymbol{W}_2\sigma(\boldsymbol{W}_1\boldsymbol{x}) \tag{2}$$

式中　$\boldsymbol{x}$、$\boldsymbol{y}$——残差块输入、输出向量

$\boldsymbol{F}(\boldsymbol{x},\{\boldsymbol{W}_i\})$——要学习的残差映射

σ——ReLU 函数

$\boldsymbol{W}_i$——权重

当 $\boldsymbol{x}$ 和 $\boldsymbol{F}$ 的维度不相同时,需要对输入 $\boldsymbol{x}$ 进行线性映射来匹配维度,即

$$\boldsymbol{y}=\boldsymbol{F}(\boldsymbol{x},\{\boldsymbol{W}_i\})+\boldsymbol{W}_s\boldsymbol{x} \tag{3}$$

式中　$\boldsymbol{W}_s$——线性映射函数

对于更深层次的网络,为了减少训练时间,将 ResNet 的瓶颈(Bottleneck)架构设计成 3 层堆栈,如图 2 所示,这 3 层分别是 1×1、3×3 和 1×1 卷积,第 1 个 1×1 卷积将 256 维的通道降到 64 维,再通过另一个 1×1 卷积将维度还原,既保持了精度,又减少了计算量。神经网络层数越多,对于原始数据的映射越多,可以得到更深层次的信息,但是模型训练时间也会越长,对应的权重文件也越大,不利于模型在移动终端的部署。

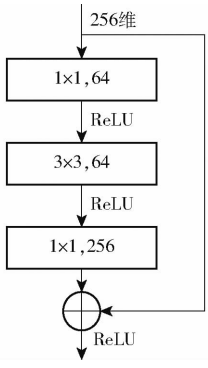


图 2　瓶颈结构

Fig. 2　Bottleneck structure

2.2　注意力机制

注意力机制最早由 MNIH 等[15]提出并引入图像分类领域,视觉注意力机制体现了人类视觉系统主动选择关注对象并加以集中处理的视觉特性,该特性能有效提升图像内容筛选、目标检索等图像处理能力。从人工智能角度看,注意力机制是机器学习中的一种数据处理方法,本质是利用相关特征图学习权重分布,再用学到的权重施加在原特征图之上,最后进行加权求和以快速提取稀疏数据的重要特征[24]。

在 Transformer attention[25]的最新版本中,注意权重被表示为 4 个注意因子(ε_1、ε_2、ε_3、ε_4)的总和,这 4 个注意因子所涉及的依赖关系的性质各不相同。ε_1 对于查询和关键内容更敏感;ε_2 更关注查询内容和相对位置;ε_3 仅仅关注关键内容;ε_4 仅仅关注相对位置。ZHU 等[26]对当前空间注意机制进行深入研究,通过分析不同注意因子的不同组合对于不同领域(图像目标检测、图像语义分割、神经机器翻译)的效果,得出注意因子为 ε_3(Key content only)的空间注意力机制,比 4 个注意因子(ε_1、ε_2、ε_3、ε_4)总和的空间注意力机制在图像方面的精度和效率更佳。

目标检测模型的 3 大组件(backbone、neck、

head)中,backbone(本文采用 ResNet 网络)的主要作用是特征提取,另外 ResNet 网络由 5 个阶段组成,其中阶段 2 ~ 5 都由瓶颈层组成,瓶颈层的主要作用是进行特征提取。因此本文将在 ResNet 网络的阶段 2 ~ 5 的瓶颈层中串联插入一个空间注意力模块,如图 3 所示,增强有用信息,抑制噪声等干扰元素的权重。并且继续探索在 ResNet 不同的阶段中加入空间注意力机制对于田间障碍物实例分割模型鲁棒性的影响。

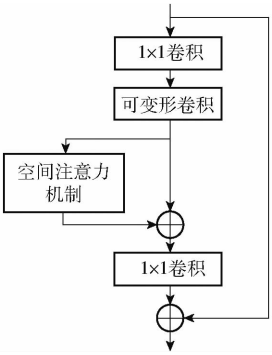


图 3 加入空间注意力机制模块的残差块结构

Fig. 3 Residual block diagram with added spatial attention mechanism

2.3 可变形卷积

由于非结构化田间障碍物形态各异,面积大小不一,这给障碍物识别任务带来了很大的困难,而且以往的卷积神经网络对整体特征的提取是依靠其固定的卷积结构,对于形态各异的目标特征提取的适应、调节能力较弱,目标识别能力不强,泛化能力差。实际上,传统的神经网络的卷积核通常是固定尺寸、固定大小(3 × 3、5 × 5),难以自适应目标的形状变化[18]。为了解决限制传统卷积神经网络识别能力的这一难题,DAI 等[19]提出了一种可变形卷积网络,替代传统的标准卷积,经研究表明,通过可变形卷积网络增加可训练的偏移量,从而适应目标形状的变化,有利于提高目标检测的鲁棒性[27-29]。

二维卷积的操作步骤为:①在输入特征图 x 上使用规则网格 $\boldsymbol{R}$ 进行采样。②用 ω 加权的采样值进行求和。一个 3 × 3 的卷积为

$$\boldsymbol{R}=\{(-1,-1),(-1,0),\cdots,(0,1),(1,1)\} \tag{4}$$

对于标准的卷积过程,输出特征图 y 中每个位置 $\boldsymbol{y}(\boldsymbol{P}_0)$ 的计算公式为

$$\boldsymbol{y}(\boldsymbol{P}_0)=\sum_{\boldsymbol{P}_n\in\boldsymbol{R}}\omega(\boldsymbol{P}_n)\boldsymbol{x}(\boldsymbol{P}_0+\boldsymbol{P}_n) \tag{5}$$

式中 $\boldsymbol{P}_n$——$\boldsymbol{R}$ 中所有采样位置

$\boldsymbol{P}_0$——输入特征图中每个位置

可变形卷积过程公式为

$$\boldsymbol{y}(\boldsymbol{P}_0)=\sum_{\boldsymbol{P}_n\in\boldsymbol{R}}\omega(\boldsymbol{P}_n)\boldsymbol{x}(\boldsymbol{P}_0+\boldsymbol{P}_n+\Delta\boldsymbol{P}_n) \tag{6}$$

式中 $\Delta\boldsymbol{P}_n$——采样点偏移量

可见,可变形卷积就是在传统的卷积操作上加入了一个采样点的偏移量 $\Delta\boldsymbol{P}_n$,以调整关键元素的采样位置,如图 4 所示。可变形卷积只为神经网络模型增加少量的参数和计算,但是大大提高了目标检测的精度[30]。

本文利用可变形卷积替代 ResNet 网络瓶颈层中的 3 × 3 卷积,与空间注意力机制共同改进 ResNet 网络,改进得到的基于空间注意力机制和可变形卷积的实例分割网络模型(ResNet - 50 + SA + DCN (2,3,5))整体结构如图 5 所示。

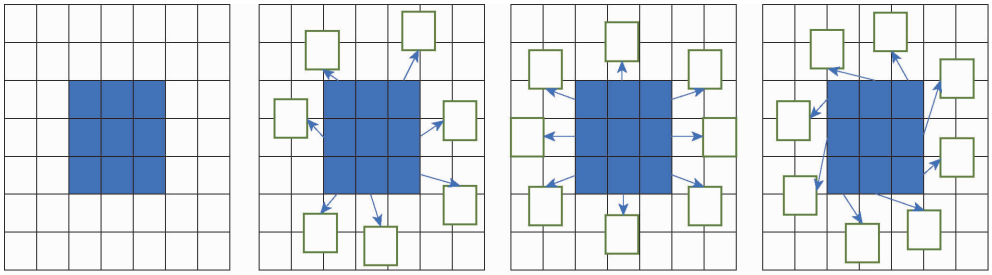
3 试验与结果分析

3.1 试验环境

试验选用的处理器为 Intel(R) Core(TM) i7 - 10700K,主频 3.8 GHz,8 核,16 MB 缓存;64 GB 内存;NVIDIA GeForce RTX2080TI(11GB) GPU 用于加速计算。操作系统是 Ubuntu 20.04,编程语言选用 Python,选择 PyTorch 深度学习框架实现网络模型的搭建、训练和调试。

3.2 模型训练与对比分析

考虑模型训练效果以及试验条件,本文模型采用迁移学习,主干网络采用 ImageNet 预训练的 ResNet - 50 网络作为初始输入权重。模型训练的周期为 24,每个周期迭代的次数为 2 400;设置学习率为 0.002 5,采用线性增加策略动态调整学习率,初



(a) 标准卷积的常规采样方式(蓝色) (b) 加入偏移量(箭头)后的卷积核(绿色) (c) 加入偏移量后卷积核的纵横变换(绿色) (d) 加入偏移量后卷积核的旋转变换(绿色)

图 4 卷积核大小为 3 × 3 的正常卷积核可变形卷积的采样方式

Fig. 4 Illustration of sampling locations in 3 × 3 standard and deformable convolutions

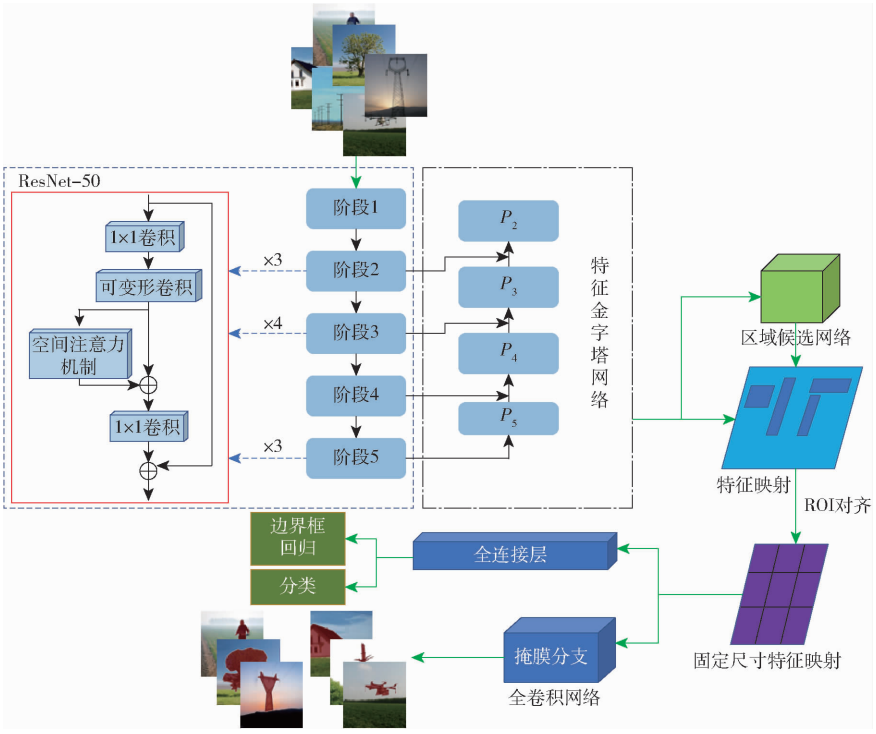


图 5　利用可变形卷积和空间注意力机制改进的 Mask R－CNN 实例分割网络

Fig. 5　Improved Mask R－CNN model with deformable convolution and spatial attention mechanism

始学习率为 2.4×10^{-4}，当迭代次数为 500 时，学习率调整为预设置的 2.5×10^{-3}；动量因子为 0.9，权重衰减系数为 0.000 1，优化算法为随机梯度下降（SGD），损失函数为对数交叉熵损失（Cross entropy loss）。

3.2.1　主干网络选择分析

Mask R－CNN 模型的主干网络选择 ResNet－50、ResNet－101，通过平均精度均值（mAP）、参数量、推断时间和损失值对比，确定适合非结构化田间障碍物实例分割的主干网络深度。试验中，控制其他条件不变，只改变主干网络的深度，两个不同深度模型的性能对比如图 6 所示。

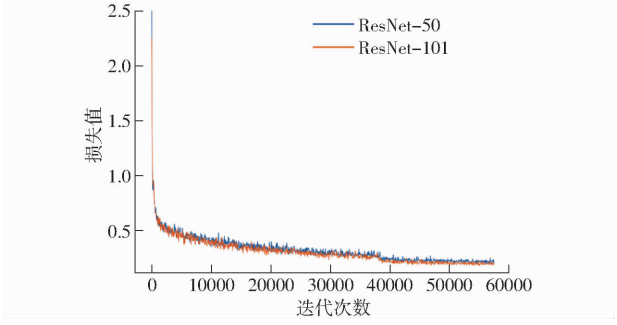


图 6　ResNet－50 和 ResNet－101 的损失值曲线

Fig. 6　Loss curves of ResNet－50 and ResNet－101

图 6 中 ResNet－50 和 ResNet－101 的损失值均随着迭代次数的增加逐渐下降并收敛，最终趋于稳定。2 个网络的损失曲线相差不大，基本重合，模型训练的总损失分别约为 0.2 和 0.18，一定程度上说明了 2 个模型具有相似的性能。此外通过表 1 的 mAP 比较，可以看出 ResNet－101 的 mAP 略微高于 ResNet－50，但是相差不大，仅为 2 个百分点左右，但是 ResNet－101 模型参数量远高于 ResNet－50，约为 6.276×10^{7}；推断时间也比 ResNet－50 长。考虑到非结构化障碍物识别模型将用于无人机，且无人机检测需要实时性强，机载端内存有限，考虑到 ResNet－101 网络对本文研究的非结构化农田障碍物分割提取任务有较大的冗余，降低网络深度对模型的性能影响不大，但是能提高模型的运算速度。综上所述，选择 ResNet－50 最为合适。

本文以 ResNet－50 为主干网络构建 Mask R－CNN 实例分割模型，并用空间注意力机制和可变形卷积对主干网络进行改进。确认模型深度为 ResNet－50 后，分析利用空间注意力机制和可变形卷积改进模型的有效性。首先利用空间注意力机制对 Mask R－CNN 进行改进，与原网络性能进行比较。主要从 mAP、AP_{50}、AP_{75}、AP_S、AP_M、AP_L、参数量和推断时间进行性能分析。增加注意力机制模型的测试结果如表 1、2 所示。表 1 中，mAP 指的是交并比从 0.5 开始，间隔 0.05 一直取值到 0.95 然后求得的平均值；AP_{50} 指交并比为 0.5 时的平均精度；AP_{75} 指交并比为 0.75 时的平均精度；AP_S、AP_M、AP_L 分别对应面积小于 32^2 像素（小目标物体）、面积大于 32^2 像素小于 96^2 像素（中等目标物体），面积大于 96^2 像素（大目标物体）测试平均精度。

表 1 不同模型的性能对比

Tab. 1 Performance comparison of different models

主干网络	Bbox						Mask						参数量	推断时间/ms
	mAP/%	AP_{50}/%	AP_{75}/%	AP_S/%	AP_M/%	AP_L/%	mAP/%	AP_{50}/%	AP_{75}/%	AP_S/%	AP_M/%	AP_L/%		
Mask R-CNN (ResNet-50)	64.5	85.8	77.6	27.8	67.9	76.7	56.9	86.3	63.0	18.9	56.2	73.6	4.377×10^7	20.3
ResNet-101	66.4	86.5	78.5	32.9	69.2	78.8	57.5	87.3	63.6	17.9	56.6	74.2	6.276×10^7	23.8
ResNet-50+SA	70.3	91.0	82.7	47.9	72.2	77.5	61.2	92.0	69.0	36.9	59.6	74.5	4.752×10^7	23.5
ResNet-50+SA+DCN (2,3)	70.4	92.3	81.9	45.7	73.1	77.7	61.2	92.7	69.3	38.1	60.6	74.3	4.405×10^7	22.8
ResNet-50+SA+DCN (2,4)	71.5	92.2	83.5	47.7	74.1	78.7	61.8	92.4	69.7	35.3	60.9	75.2	4.520×10^7	23.9
ResNet-50+SA+DCN (2,5)	70.8	91.9	82.9	47.6	73.1	78.0	62.2	92.9	70.9	37.1	61.5	74.9	4.638×10^7	22.8
ResNet-50+SA+DCN (3,4)	71.5	91.7	83.1	48.7	73.3	79.1	62.1	92.5	70.1	35.8	61.2	75.0	4.549×10^7	24.5
ResNet-50+SA+DCN (3,5)	71.4	92.0	84.0	49.3	73.8	78.5	61.7	92.9	70.6	37.5	60.2	75.2	4.666×10^7	22.4
ResNet-50+SA+DCN (4,5)	71.9	92.2	82.6	49.7	73.9	79.6	62.0	92.4	70.7	37.5	61.1	74.9	4.782×10^7	23.1
ResNet-50+SA+DCN (2,3,4)	71.3	91.5	81.7	44.1	73.2	79.3	61.9	92.4	69.9	35.1	61.0	74.6	4.549×10^7	25.2
ResNet-50+SA+DCN (2,3,5)	71.3	91.6	82.7	48.7	73.2	79.1	62.3	92.6	71.9	38.5	61.5	74.7	4.666×10^7	22.9
ResNet-50+SA+DCN (3,4,5)	71.4	92.4	82.3	49.4	74.4	78.2	61.8	92.7	69.2	36.3	60.8	74.8	4.810×10^7	24.7
ResNet-50+SA+DCN (2,3,4,5)	71.7	92.1	82.5	48.0	74.4	78.9	61.9	92.7	70.0	37.4	61.5	74.3	4.810×10^7	26.1
PointRend	69.8	92.4	81.6	45.8	72.0	76.4	65.8	92.8	77.3	35.7	65.1	79.9	5.576×10^7	26.1
SOLO							47.0	83.3	48.6	13.0	42.4	64.2	3.610×10^7	24.2
YOLACT	65.4	91.4	75.5	48.5	64.9	73.9	60.0	88.7	68.5	23.6	60.5	76.3	5.380×10^7	27.3

表 2 不同模型各个类别的 AP 值对比

Tab. 2 Performance comparison of different models in each category %

主干网络	Bbox						Mask					
	无人机	建筑	电线塔	人	树木	电线杆	无人机	建筑	电线塔	人	树木	电线杆
Mask R-CNN (ResNet-50)	81.0	63.1	64.6	65.5	72.6	40.1	51.8	67.3	57.6	67.8	69.3	27.5
ResNet-50+SA	82.5	66.1	76.7	69.7	73.4	53.6	56.3	70.9	66.9	70.0	70.1	33.1
ResNet-50+SA+DCN (2,3)	82.9	65.3	77.1	68.9	75.5	52.6	55.9	70.3	66.3	70.1	72.1	32.6
ResNet-50+SA+DCN (2,4)	84.7	68.4	77.1	69.3	75.7	53.8	56.8	72.3	65.8	70.3	72.0	33.6
ResNet-50+SA+DCN (2,5)	82.5	67.4	77.5	68.9	75.1	53.2	57.2	73.3	66.8	69.7	72.5	33.7
ResNet-50+SA+DCN (3,4)	84.0	66.8	76.1	72.6	75.3	54.0	58.0	71.0	66.6	69.7	73.2	34.1
ResNet-50+SA+DCN (3,5)	84.0	69.8	76.8	69.6	75.6	52.6	56.8	72.4	66.1	70.4	71.1	33.4
ResNet-50+SA+DCN (4,5)	85.1	69.0	77.4	70.2	75.7	54.1	57.6	72.7	67.0	69.9	71.6	33.4
ResNet-50+SA+DCN (2,3,4)	83.8	68.5	76.5	71.2	74.8	53.0	57.6	72.1	66.4	70.5	71.3	33.9
ResNet-50+SA+DCN (2,3,5)	83.9	68.4	76.6	69.6	75.2	54.0	58.0	72.2	66.7	70.6	71.6	34.8
ResNet-50+SA+DCN (3,4,5)	84.9	65.7	77.7	70.6	75.4	54.0	57.8	69.6	66.5	70.9	72.5	33.3
ResNet-50+SA+DCN (2,3,4,5)	84.4	67.2	77.6	68.9	77.2	54.8	57.7	69.6	66.9	70.9	72.3	34.2
PointRend	82.7	66.2	74.9	67.7	75.1	52.0	66.7	72.3	70.5	72.2	73.8	39.1
SOLO							44.2	49.1	57.3	54.9	55.1	21.4
YOLACT	79.1	62.2	69.7	65.4	70.8	44.9	62.1	69.4	67.9	67.1	70.8	22.8

3.2.2 改进后的网络性能分析

由表 1 可知,加入空间注意力机制后的模型(ResNet-50+SA)比原模型在各项性能上都有了不同程度的提升。从 Bbox 来看,ResNet-50+SA 模型比改进前模型的 mAP 高 5.8 个百分点,特别是小面积物体的平均精度(AP_S),提高 20.1 个百分点;从 Mask 来看,改进后比改进前模型的 mAP 提高 4.3 个百分点,AP_S提升比较显著,为 18 个百分点;另外改进后模型的参数量仅增加 8.6%。

从表 2 可知,不论是 Bbox 还是 Mask,加入空间注意力机制后的模型比 Mask R-CNN 模型性能都有提升,其中电线杆的特征是细长,属于小面积目标。这种小面积目标的平均精度(AP)从 40.1%、27.5%提升到 53.6%、33.1%,分别提高 13.5、5.6 个百分点。

从 AP_S和电线杆 AP 可知,空间注意力机制提高了模型对于细小物体特征的提取能力。空间注意力机制的引入可以在获得较高 AP 值的基础上,使模型参数量增长较少。

在加入空间注意力机制的基础上,将瓶颈层的 3×3 卷积调整为可变形卷积,两者结合共同改进 Mask R-CNN,为了进一步优化 ResNet-50+SA +

DCN 模型的性能，本文从 ResNet 阶段 2～5 的组合((2,3)、(2,4)、(2,5)、(3,4)、(3,5)、(4,5)、(2,3,4)、(2,3,5)、(3,4,5)、(2,3,4,5))中分别引入 2 个模块，并对这些组合进行遍历，分析试验在不同阶段组合中引入空间注意力模块和可变形卷积模块对于模型的影响。其中，(2,3)是从 ResNet 的阶段 2、阶段 3 引入 2 个模块；(3,4,5)是从 ResNet 的阶段 3、阶段 4、阶段 5 引入 2 个模块；(2,3,4,5)是从 ResNet 的阶段 2、阶段 3、阶段 4、阶段 5 引入 2 个模块，以此类推。测试结果如表 1、2 所示。

从表 1、2 可知，不论从 ResNet 的哪个阶段引入可变形卷积，ResNet－50＋SA＋DCN 模型的综合性能都比仅引入空间注意力机制的 ResNet－50＋SA 模型性能更优。从 ResNet 的 2 个阶段引入空间注意力机制和可变形卷积分析，ResNet－50＋SA＋DCN (4,5)模型比其他模型的权重文件更大；从 ResNet 的 3 个阶段引入空间注意力机制和可变形卷积分析，ResNet－50＋SA＋DCN (3,4,5)模型比其他模型的权重文件更大。由此可知，在 ResNet 的前阶段引入空间注意力机制和可变形卷积，能够加强网络前阶段对重点特征信息的提取能力和提升网络对不同尺寸的目标适应能力，减少网络后阶段需要处理的数据量，从而减少模型的参数量。从 ResNet 的全部 4 个阶段进行改进对网络性能提升不大，但是 2 个模块的引入带来的参数量会增加网络的负荷，导致模型的参数量偏大。

从总体的 mAP 值和各类别的 AP 值、参数量、推断时间进行分析，由表 1、2 可知，从 3 个阶段((3,4,5)、(2,3,5))引入空间注意力机制和可变形卷积模块比其他模型的综合性能更优。此外对比 ResNet－50＋SA＋DCN (2,3,5)和 ResNet－50＋SA＋DCN (3,4,5)两个模型，ResNet－50＋SA＋DCN (2,3,5)模型的参数量更小，速度更快；而且 ResNet－50＋SA＋DCN (2,3,5)模型在 Mask 上的 mAP 更高，而且模型的参数量比 ResNet－50 仅增长 6.6%，比 ResNet－50＋SA 的参数量更少。

由表 1 分析可知，从 Bbox 来看，本文提出的 ResNet－50＋SA＋DCN (2,3,5) 模型在 mAP 上比 YOLACT 高 5.9 个百分点，比 PointRend 高 1.5 个百分点。从 Mask 来看，ResNet－50＋SA＋DCN (2,3,5)在 mAP 上比 YOLACT 高 2.3 个百分点，比 SOLO 高 15.3 个百分点，但是比 PointRend 低 3.5 个百分点；另外，ResNet－50＋SA＋DCN (2,3,5)的 AP_S 比 YOLACT 高 14.9 个百分点，比 SOLO 高 25.5 个百分点，比 PointRend 高 2.8 个百分点。从参数量来看，SOLO 模型的参数量最少，比 ResNet－50＋SA＋DCN (2,3,5)少 1.056×10^7，其中 PointRend 的参数量最多，比 ResNet－50＋SA＋DCN (2,3,5)多 9.10×10^6；在推断时间方面，ResNet－50＋SA＋DCN (2,3,5)的推断时间比 YOLACT 少 4.4 ms，比 SOLO 少 1.3 ms，比 PointRend 少 3.2 ms。

从表 2 分析可知，从 Bbox 的 AP 值来看，ResNet－50＋SA＋DCN (2,3,5) 比 PointRend、YOLACT、SOLO 模型的性能都更加优异，但是从 Mask 的 AP 值来看，PointRend 模型的性能更为优异。PointRend 模型利用计算机图像渲染技术提高了 Mask 的AP 值，但是在参数量、推断时间、Bbox 方面的性能低于本文提出的 ResNet－50＋SA＋DCN (2,3,5)模型。

综上所述，本文将在 ResNet 的阶段 2、阶段 3、阶段 5 引入空间注意力机制模块和可变形卷积模块，构建适用于非结构化农田障碍物识别模型 ResNet－50＋SA＋DCN (2,3,5)，模型资源开销低，为后期将目标识别与分割模型移入内存有限的无人机设备提供了可能。

3.2.3　不同模型输出结果分析

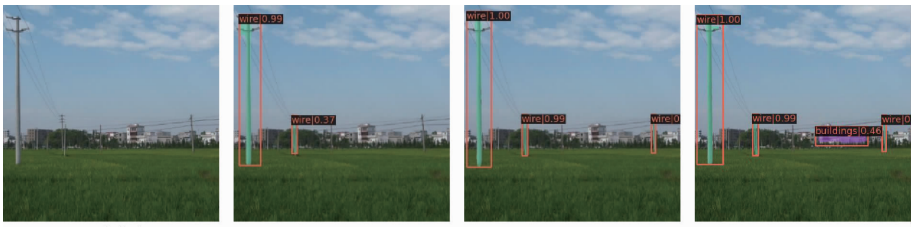
为了更直观地解释引入空间注意力机制和可变形卷积对 Mask R－CNN 模型性能的提升，通过图像输出结果对不同模型进行分析。

如图 7b 所示，目标人物的头部没有完全识别，加入空间注意力机制后，模型识别效果(图 7c、7d)比 Mask R－CNN 模型的识别(图 7b)更准确；其次图 7c 目标手部还没有完全覆盖，加入可变形卷积模块后(图 7d)，目标的轮廓分割效果最好，网络提取的特征更好地覆盖在目标对象区域。如图 8c、8d 所示，加入空间注意力模块后的模型特征提取能力更强，可以识别到更多的物体。此外，加入可变形卷积

(a) 原图　(b) ResNet-50　(c) ResNet-50+SA　(d) ResNet-50+SA+DCN

图 7　不同模型的输出结果(人)

Fig. 7　Output results of different models(person)



(a) 原图 (b) ResNet-50 (c) ResNet-50+SA (d) ResNet-50+SA+DCN

图 8 不同模型的输出结果(电线杆与建筑)

Fig. 8 Output results of different models(wire pole and building)

后模型 ResNet－50 + SA + DCN 能够识别到更远处的物体(图 8d)。但是对于较远处被阻挡的目标还是会出现无法识别的情况。

综上所述,ResNet－50 + SA 和 ResNet－50 + SA + DCN 模型提取的特征更好地覆盖在目标对象区域,说明在现有的 Mask R－CNN 模型加入空间注意力机制模块可以增强有用信息,提高网络模型的特征提取能力;而加入可变形卷积模块可以使模型增大感受野,提高模型对目标不同尺寸的适应能力,进一步说明了本文对网络的改进是有效的,可以为无人机障碍物识别提供视觉支持。

4 结论

(1)为了建立适用于无人机田间障碍物识别的网络模型,本文以空间注意力机制和可变形卷积改进 ResNet 特征提取网络,进而优化 Mask R－CNN 实例分割模型,可以实现非结构化农田障碍物的识别与分割。

(2)为了提高利用空间注意力机制和可变形卷积改进 Mask R－CNN 模型的有效性,分析从 ResNet 的阶段 2～5 中的不同组合中引入 2 个模块对于模型性能的影响,最终确定在 ResNet 的阶段 2、阶段 3、阶段 5 引入 2 个模块的性能最优,ResNet－50 + SA + DCN(2,3,5)模型的 Bbox 和 Mask 的 mAP 值分别为 71.3%、62.3%,与仅加入空间注意力机制的模型相比,Bbox 和 Mask 的 mAP 值分别提高 1.0、1.1 个百分点,参数量和推断时间也有了相应的优化。

(3)与 YOLACT、SOLO、PointRend 模型相比,ResNet－50 + SA + DCN (2,3,5)在 Bbox 上的 mAP 更高,推断时间更短,实时性更好;另外,与 Mask R－CNN 模型相比,本文模型在小面积目标的检测方面,性能更加优异。因此,ResNet－50 + SA + DCN (2,3,5)模型在非结构化田间障碍物识别与分割任务中具有优异的表现,并且在控制模型检测速度的情况下,用很小的资源开销明显提升了模型检测准确率,较好地平衡了模型复杂度和识别精度,充分证明了 ResNet－50 + SA + DCN (2,3,5)模型在非结构化农田障碍物识别与分割上的优越性。

参 考 文 献

[1] 周志艳,明锐,臧禹,等. 中国农业航空发展现状及对策建议[J]. 农业工程学报, 2017, 33(20):1－13.
ZHOU Zhiyan, MING Rui, ZANG Yu, et al. Development status and countermeasures of agricultural aviation in China[J]. Transactions of the CSAE, 2017, 33(20): 1－13. (in Chinese)

[2] 兰玉彬,王林琳,张亚莉. 农用无人机避障技术的应用现状及展望[J]. 农业工程学报, 2018, 34(9):104－113.
LAN Yubin, WANG Linlin, ZHANG Yali. Application and prospect on obstacle avoidance technology for agricultural UAV[J]. Transactions of the CSAE, 2018, 34(9): 104－113. (in Chinese)

[3] MOFFATT A, PLATT E, MONDRAGON B, et al. Obstacle detection and avoidance system for small UAVs using a LiDAR[C]// 2020 International Conference on Unmanned Aircraft Systems, 2020: 633－640.

[4] LUDENO G, CATAPANO I, RENGA A, et al. Assessment of a micro-UAV system for microwave tomography radar imaging [J]. Remote Sensing of Environment, 2018, 212: 90－102.

[5] XU Yuefan, ZHU Minling, XU Yue, et al. Design and implementation of UAV obstacle avoidance system[C]// 2019 2nd International Conference on Safety Produce Informatization, 2019: 275－278.

[6] WANG Y, YANG C, HU M, et al. Identification of deep breath while moving forward based on multiple body regions and graph signal analysis[C]// IEEE International Conference on Acoustics, Speech and Signal Processing, 2021: 7958－7962.

[7] JI Yuhan, LI Shichao, PENG Cheng, et al. Obstacle detection and recognition in farmland based on fusion point cloud data [J]. Computers and Electronics in Agriculture, 2021, 189: 106409.

[8] 傅隆生, 冯亚利, ELKAMIL T, 等. 基于卷积神经网络的田间多簇猕猴桃图像识别方法[J]. 农业工程学报, 2018, 34(2): 205－211.
FU Longsheng, FENG Yali, ELKAMIL T, et al. Image recognition method of multi-cluster kiwifruit in field based on convolutional neural networks[J]. Transactions of the CSAE, 2018, 34(2): 205－211. (in Chinese)

[9] 刘芳, 刘玉坤, 林森,等. 基于改进型 YOLO 的复杂环境下番茄果实快速识别方法[J]. 农业机械学报, 2020, 51(6):

229 – 237.
LIU Fang, LIU Yukun, LIN Sen, et al. Fast recognition method for tomatoes under complex environments based on improved YOLO[J]. Transactions of the Chinese Society for Agricultural Machinery, 2020, 51(6): 229 – 237. (in Chinese)

[10] 孙哲, 张春龙, 葛鲁镇,等. 基于 Faster R – CNN 的田间西兰花幼苗图像检测方法[J]. 农业机械学报, 2019, 50(7): 216 – 227.
SUN Zhe, ZHANG Chunlong, GE Luzhen, et al. Image detection method for broccoli seedlings in field based on Faster R – CNN[J]. Transactions of the Chinese Society for Agricultural Machinery, 2019, 50(7): 216 – 227. (in Chinese)

[11] HE K, GKIOXARI G, DOLLAR P, et al. Mask R – CNN[C]// Proceedings of the IEEE International Conference on Computer Vision, 2017: 2961 – 2969.

[12] JIA Weikuan, TIAN Yuyu, LUO Rong, et al. Detection and segmentation of overlapped fruits based on optimized Mask R – CNN application in apple harvesting robot[J]. Computers and Electronics in Agriculture, 2020, 172:105380.

[13] TIAN Yunong, YANG Guodong, WANG Zhe, et al. Instance segmentation of apple flowers using the improved Mask R – CNN model[J]. Biosystems Engineering, 2020, 193: 264 – 278.

[14] YU Yang, ZHANG Kailiang, LI Yang, et al. Fruit detection for strawberry harvesting robot in non-structural environment based on Mask-RCNN[J]. Computers and Electronics in Agriculture, 2019, 163:104846.

[15] MNIH V, HEESS N, GRAVES A. Recurrent models of visual attention[C]// Advances in Neural Information Processing Systems, 2014: 2204 – 2212.

[16] 黄林生, 罗耀武, 杨小冬,等. 基于注意力机制和多尺度残差网络的农作物病害识别[J]. 农业机械学报, 2021, 52(10): 264 – 271.
HUANG Linsheng, LUO Yaowu, YANG Xiaodong, et al. Crop disease recognition based on attention mechanism and multi-scale residual network[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(10): 264 – 271. (in Chinese)

[17] 熊俊涛, 刘柏林, 钟灼,等. 基于深度语义分割网络的荔枝花叶分割与识别[J]. 农业机械学报, 2021, 52(6): 252 – 258.
XIONG Juntao, LIU Bolin, ZHONG Zhuo, et al. Litchi flower and leaf segmentation and recognition based on deep semantic segmentation[J]. Transactions of the Chinese Society for Agricultural Machinery, 2021, 52(6): 252 – 258. (in Chinese)

[18] JADERBERG M, SIMONYAN K, ZISSERMAN A. Spatial transformer networks[J]. Advances in Neural Information Processing Systems, 2015, 28: 2017 – 2025.

[19] DAI Jifeng, QI Haozhi, XIONG Yuwen, et al. Deformable convolutional networks[C]// Proceedings of the IEEE International Conference on Computer Vision, 2017: 764 – 773.

[20] SUN Jun, YANG Kaifeng, HE Xiaofei, et al. Beet seedling and weed recognition based on convolutional neural network and multi-modality images[J]. Multimedia Tools and Applications, 2022, 81: 5239 – 5258.

[21] WANG Dashuai, LI Wei, LIU Xiaoguang, et al. UAV environmental perception and autonomous obstacle avoidance: a deep learning and depth camera combined solution[J]. Computers and Electronics in Agriculture, 2020, 175:105523.

[22] BENGIO Y, SIMARD P, FRASCONI P. Learning long-term dependencies with gradient descent is difficult[J]. IEEE Transactions on Neural Networks, 1994, 5(2):157 – 166.

[23] HE Kaiming, ZHANG Xiangyu, REN Shaoqing, et al. Deep residual learning for image recognition[C]// Proceedings of the IEEE Conference on Computer Vision and Pattern Recognition, 2016: 770 – 778.

[24] CHAUDHARI S, MITHAL V, POLATKAN G, et al. An attentive survey of attention models[J]. ACM Transactions on Intelligent Systems and Technology, 2021, 12(5): 1 – 32.

[25] DAI Zihang, YANG Zhilin, YANG Yiming, et al. Transformer-XL: attentive language models beyond a fixed-length context[J]. arXiv preprint arXiv:1901.02860, 2019.

[26] ZHU Xizhou, CHENG Dazhi, ZHANG Zheng, et al. An empirical study of spatial attention mechanisms in deep networks[C]// Proceedings of the IEEE/CVF International Conference on Computer Vision,2019: 6688 – 6697.

[27] ZHANG Chen, KIM J. Object detection with location-aware deformable convolution and backward attention filtering[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 9452 – 9461.

[28] DENG Liuyuan, YANG Ming, LI Hao, et al. Restricted deformable convolution-based road scene semantic segmentation using surround view cameras[J]. IEEE Transactions on Intelligent Transportation Systems, 2019, 21(10): 4350 – 4362.

[29] LIU Zhenyu, YANG Benyi, DUAN Guifang, et al. Visual defect inspection of metal part surface via deformable convolution and concatenate feature pyramid neural networks[J]. IEEE Transactions on Instrumentation and Measurement, 2020, 69(12): 9681 – 9684.

[30] ZHU Xizhou, HU Han, LIN S, et al. Deformable convnets v2: more deformable, better results[C]// Proceedings of the IEEE/CVF Conference on Computer Vision and Pattern Recognition, 2019: 9308 – 9316.